The Conventional versus the Proposed Paradigm for Incremental Sampling of Soils at Military Ranges

Presenter

Thomas Georgian, Ph.D., Chemist

Environmental and Munitions Center of Expertise

US Army Engineering and Support Center, Huntsville

EMDQ Workshop

26 – 29 March 2012

US Army Corps of Engineers

BUILDING STRONG®

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Conventional versus the Proposed Paradigm for Incremental Sampling of Soils at Military Ranges	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **US Army Engineering and Support Center, Huntsville,Environmental and Munitions Center of Expertise,4820 University Square,Huntsville,AL,35816-1822**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 9th Annual DoD Environmental Monitoring and Data Quality (EDMQ) Workshop Held 26-29 March 2012 in La Jolla, CA.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**15**	

Conventional / Classic Incremental Sampling Design Paradigm (Strategy)

- Collect *n* = 1 incremental sample (IS) for each Sampling Unit (SU) & field triplicates for 10% of SUs.
- If %RSD < 30%, results fully usable. Otherwise, flag results (e.g., as estimated).
- %RSDs of SUs with triplicates "represent" precision of SUs for which *n* = 1.
- Decisions based on one IS result acceptable (e.g., maximum of SU is compared with a decision limit).

"Conventional" Approach

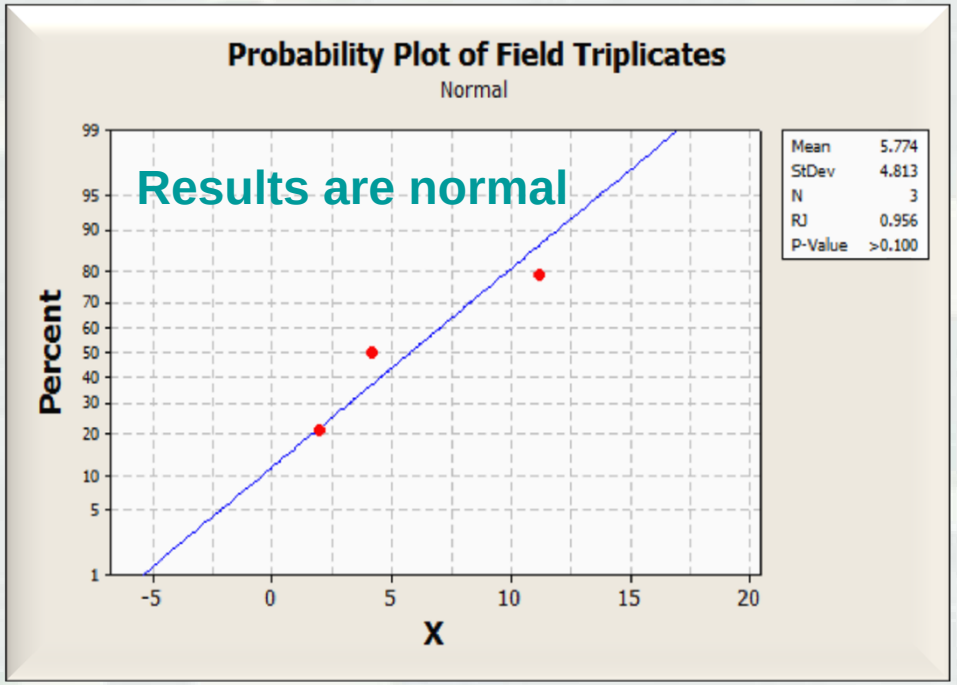
- Advantages: Cost effective and simple to understand and implement.

- Disadvantage: Data of unknown quality; ***decision errors*** are not controlled.

  - ***False positive (FP) error, α***
    (e.g., wrongly concluding
    site conc. > background/action level)
  - ***False negative (FN) error, β***
    (e.g., wrongly concluding
    site conc. ≤ background/action level)

Example of Classic Approach

- IS Field triplicates:
  - 4.2
  - 2.0
  - 11.2



Sample Statistics

$\bar{x} = 5.8, \; s = 4.8$

$\%RSD = (s / \bar{x}) \times 100$

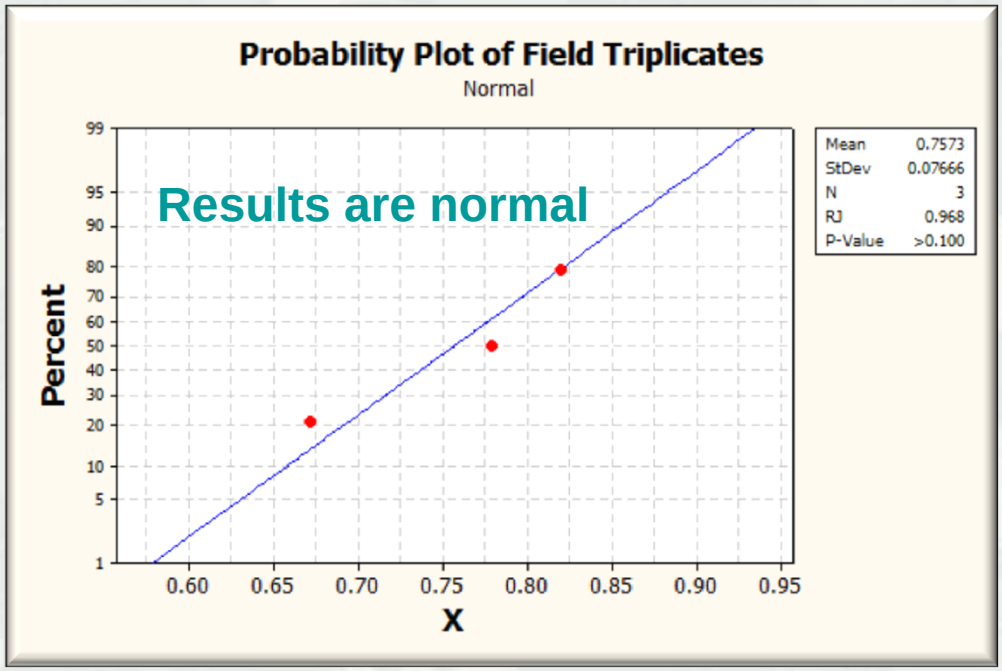
$= \mathbf{83\%} >> 30\%$

s = sample standard deviation

$\bar{x}$ = sample mean

Re-sampling of Triplicates

- Field triplicates:
  - 0.82
  - 0.67
  - 0.78

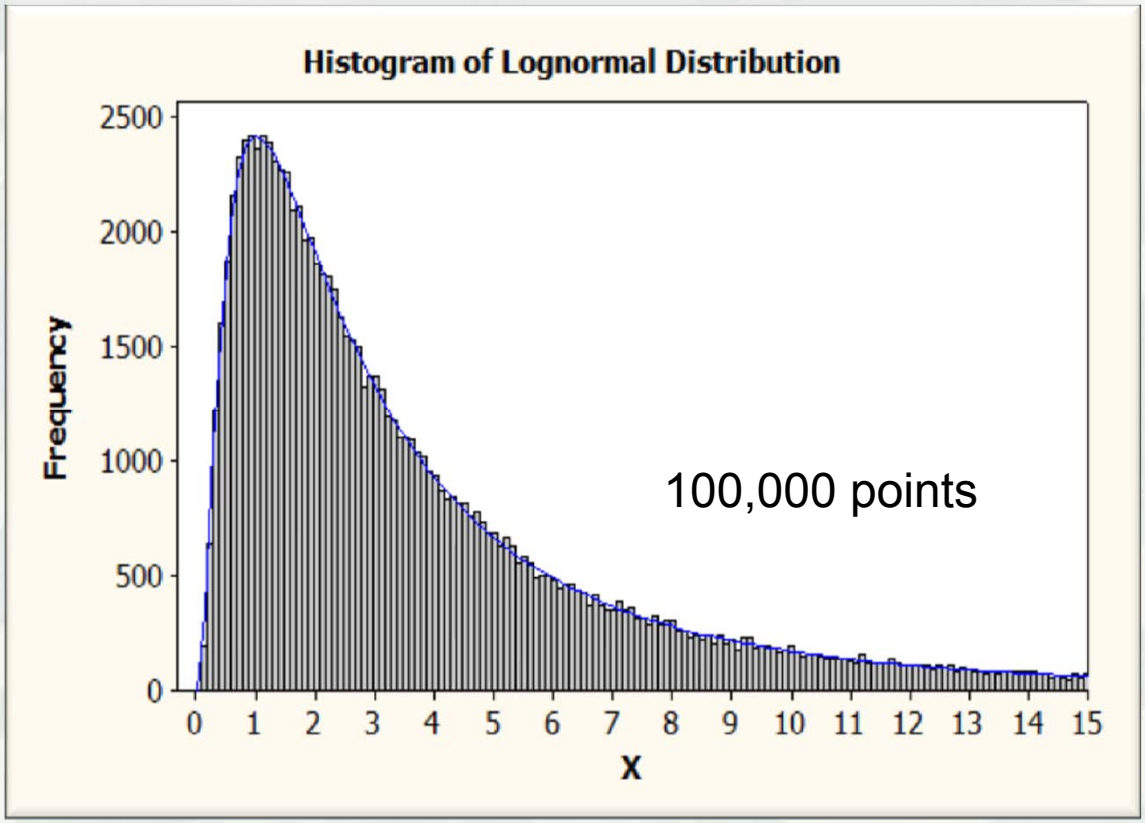


Sample Statistics

$\bar{x} = 0.76, \; s = 0.077$

%RSD = **10%** < 30%

- The two sets of triplicates were actually randomly selected from a lognormal distribution.



Population Parameters

$\mu \approx 4.5, \ \sigma \approx 5.9$

%RSD = 130%

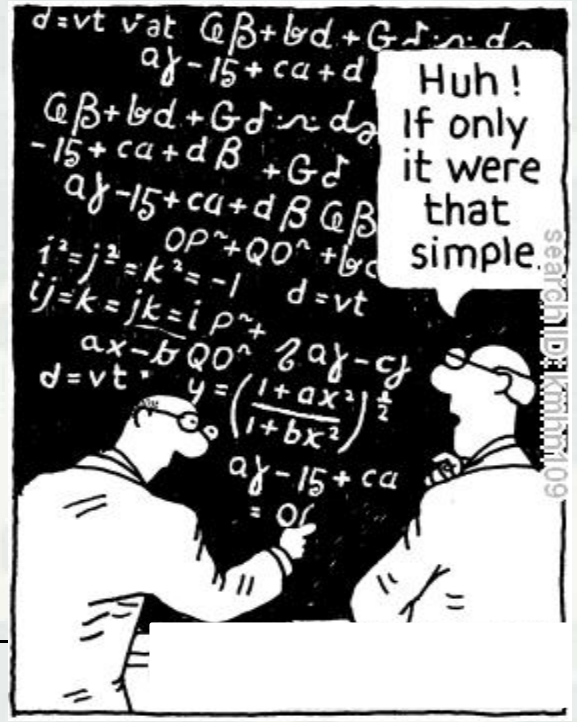
Sample %RSD	$\bar{x}$	95%UCL
83%	5 .8	14
10%	0.76	0.89

- Distribution cannot be reliably determined when *n* = 3.

ITRC Guidance Incremental Sampling Methodology (ISM) *

...the RSD does not provide an indication of the accuracy of the estimate of the mean or 95% UCL...

http://www.itrcweb.org/ISM-1/Executive_Summary.html



* ISM-1, Interstate Technology Regulatory Council, February 2012

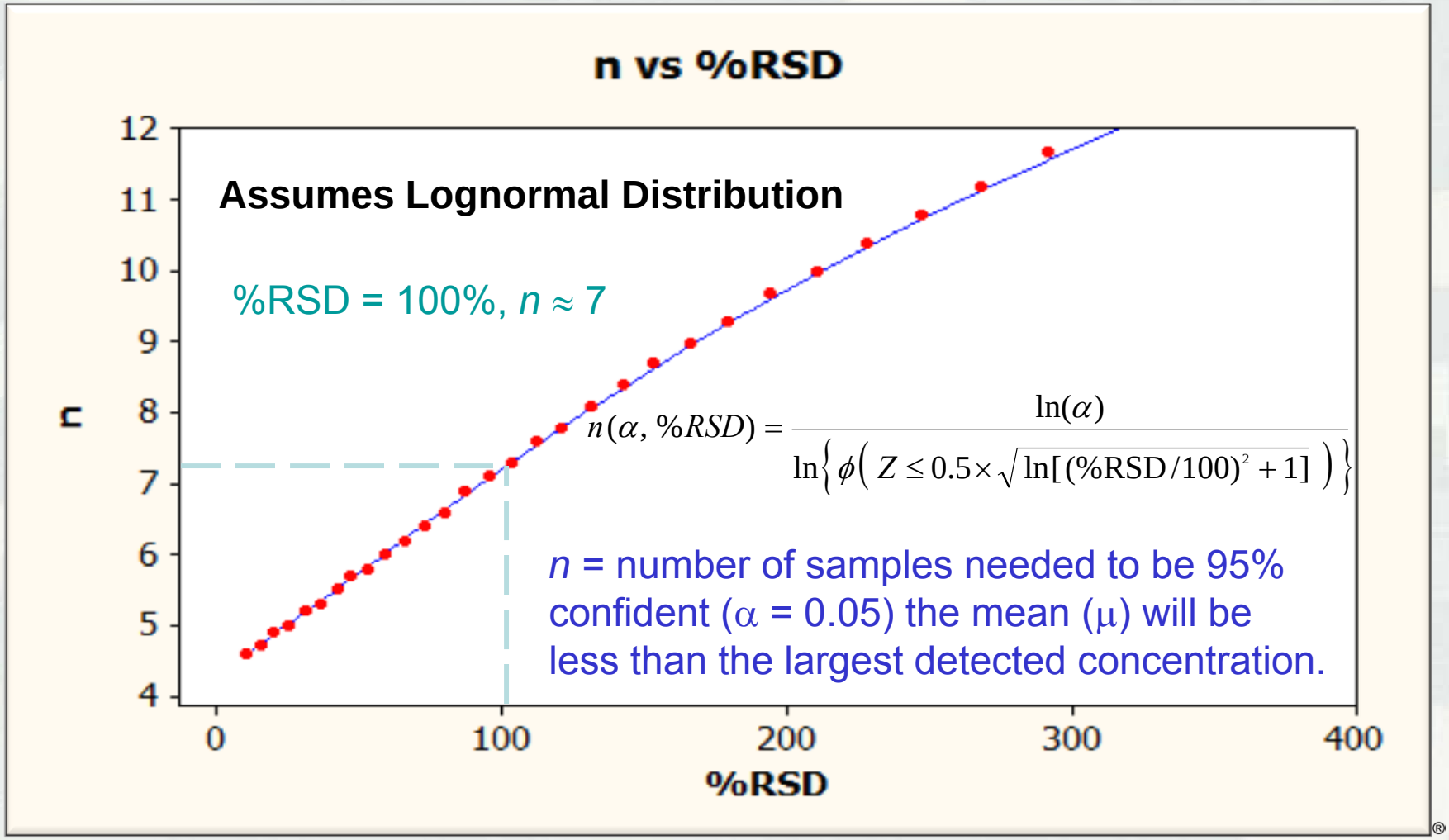
ITRC ISM Guidance

When a site includes many [Decision Units] DUs, it may be tempting to extrapolate the estimate of the variance (or the CV) [RSD] from one DU to another…

…Statistical theory suggests that we can expect the estimated mean and SD [standard deviation] to be independent for normal distributions but positively correlated for positively skewed distributions…If …independent,…high estimated means may have low SDs and vice versa…By contrast, if the parameters are correlated because of some asymmetry in the distribution…it would be preferable to extrapolate the average CV…A priori knowledge about the distribution shape is unlikely…

Translation: Don't extrapolate results; it doesn't work!

Comparing the maximum with an Action Level (AL)



Revised/Proposed Paradigm

- Use Data Quality Objective (DQO) or Technical Project Planning (TPP) Process.
- Determine replicates *n* for each SU using statistical approach (e.g., Visual Sampling Plan).

$$n \uparrow \rightarrow \alpha, \beta, (\Delta/\sigma) \downarrow$$

α = *FP error,* β = *FN error (e.g.,* $\alpha, \beta = 0.1 - 0.01$ *)*

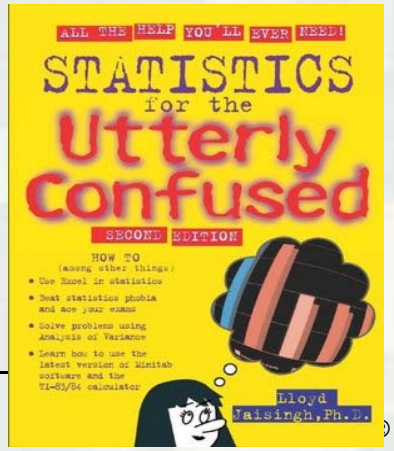
Δ = *Change important to detect (e.g.,* $\Delta = AL - \mu$*)*

σ = *Standard deviation* (Note: %RSD $\propto \sigma$)

- Use ***hypothesis tests*** to compare site means/medians with Action Level (AL) or background (BG) (e.g., using ProUCL).

  - USACE Engineer Manual (EM) 1110-1-1014, *Environmental Statistics* (Jan 2008).
  - EPA QA/G-9S, *Data Quality Assessment: Statistical Methods for Practitioners* (Feb 2006).

- Everything you wanted to know about hypothesis tests but were afraid to ask.



Alternatives to Hypothesis Tests

- Compare ***upper confidence limit (UCL)*** of mean with AL; conclude site is "clean" if UCL < AL.

$$\text{Student's t } 95\%\text{UCL} = \bar{x} + t_{n-1,0.95} \times s/\sqrt{n} \approx \bar{x} + (2-3) \times s/\sqrt{n}$$

(x is normal, $n \geq 3$)

$$\text{Chebyshev 95\% UCL} = \bar{x} + \sqrt{\left(\frac{1}{0.05} - 1\right)} \times s/\sqrt{n} \approx \bar{x} + 4\ s/\sqrt{n}$$

(x is not normal)

- Calculate (1-α)100% (e.g., 95%) BG ***upper prediction limit (UPL)*** to control FPs. SUs are consistent with BG if:

(Normality assumed)

$$\bar{x}_i < (1-\alpha)100\%\ \text{UPL}_{BG} = \bar{x}_{BG} + t_{1-(\alpha/k),n-1}\sqrt{\frac{1}{m}+\frac{1}{n}}\ s_{BG}$$

α	Site-wide FP error
n	Number of background samples ($n \geq 2$)
k	Number of site SUs ($k \geq 1$)
m	Number of (independent) replicates per SU ($m \geq 1$)
$\bar{x}_i$	Mean of m replicates for i^{th} SU, $i = 1, 2, \ldots, k$
$\bar{x}_{BG}$	BG mean
s_{BG}	BG standard deviation

BG sample maximum = γ 100% UPL (Non-parametric)

$\gamma = \gamma$ (n BG samples, k SUs, m replicates per SU)

$$\gamma = \left(\frac{n}{km+n}\right)\left\{\frac{\left(C_{m-1}^{m}\right)^{k}}{C_{km+n-1-k}^{km+n-1}} + C_1^k \frac{\left(C_{m-1}^{m}\right)^{k-1}}{C_{km+n-(k-1)}^{km+n-1}} + C_2^k \frac{\left(C_{m-1}^{m}\right)^{k-2}}{C_{km+n-(k-2)}^{km+n-1}} + \ldots + C_{k-1}^k \frac{C_{m-1}^{m}}{C_{km+n-2}^{km+n-1}} + 1\right\}$$

$$C_s^r = \frac{r!}{s!\,(r-s)!}, \quad s \le r$$

γ = Probability at least m-1 out of m replicates will be less than the maximum of n BG samples for each of the k SUs.

$\gamma \ge 0.9$		
	m = 2	**m = 3**
k	**n**	**n**
1	3	6
2	5	8
3	6	10
4	7	12
5	8	13

Gibbons, R. D, Some Additional Nonparametric Prediction Limits for Ground-Water Detection Monitoring at Waste Disposal Facilities, *Groundwater,* Vol. 29, No. 5 (Sep – Oct 1991)

Conventional Versus Proposed Approach

Approach	Conventional	Proposed
%RSD measures data quality	✓	×
Extrapolating %RSDs	✓	×
Decisions based one sample	✓	×
Statistical sampling design	×	✓

✓ Potentially acceptable
× Potentially unacceptable